

Appl. No. : 10/067,550  
Filed : February 5, 2002

### REMARKS

By this paper, Claim 6 is amended. Claims 1-5 and 7-35 are unchanged. New dependent Claims 36-40 are added. After entry of the foregoing amendments, Claims 1-40 are pending in the application. Reconsideration and allowance of the claims in light of the present remarks is respectfully requested. Claim 6 is amended to correct a typographical error, and is not amended to narrow the claim or to overcome any art.

#### I. Discussion of the 35 U.S.C. § 102(e) Rejections

Claims 10, 13, 15, 16, 18, 20, 22, 23, 27, 30, 32, and 33 were rejected under 35 U.S.C. § 102(e) as being anticipated by Chow et al. (U.S. Patent No. 5,692,104).

##### A. The Law of Anticipation

Anticipation under Section 102 can be found only if a reference shows exactly what is claimed. *Titanium Metals Corp. v. Banner*, 778 F. 2d 775 (Fed. Cir. 1985). More particularly, a finding of anticipation requires the disclosure in a single piece of prior art of each and every limitation of a claimed invention. *Electro Med. Sys. S.A. v. Cooper Life Sciences*, 34 F.3d 1048, 1052 (Fed. Cir. 1994). "To anticipate, every element and limitation of the claimed invention must be found in a single prior art reference, arranged as in the claim." *Brown v. 3M*, 265 F.3d 1349 (Fed. Cir. 2001). "All words in a claim must be considered in judging the patentability of that claim against the prior art." *In re Wilson*, 424 F.2d 1382, 1385 (CCPA 1970).

##### B. Discussion of Distinctions of the Claims 10, 13, 15, 16, 18, 20, 22, 23, 27, 30, 32, and 33 in view of Chow

The Chow reference is directed to detecting speech activity for a continuous speech recognition system. As shown in Figure 2, the speech recognition system in Chow includes a feature extraction process 210 for detecting speech activity. The feature extraction process receives digitized sound waveform, finds beginning and end points of the waveform, and produces quantized spectral representation vectors for the sound. Then, the vector quantization (VQ) block of the feature extraction process performs sound classification to determine whether the sound

waveform is speech or noise. (Col. 7, ll. 30-32). The VQ does this by computing the “distortion between the input spectral representation vector, corresponding to a frame of the sound sampling, and two code books, one for speech and one for noise.” See Col. 12. ll. 38-40. If the sound waveform is classified as speech, then the speech waveform is permitted to proceed to the speech recognition stage. On the other hand, if the sound waveform is classified as noise, then the sound waveform is not permitted to proceed to the speech recognition system. The speech recognition system then receives the data representing speech and performs word recognition and outputs a string of words, *i.e.*, text.

Applicant submits that Chow neither teaches nor suggests the following two features that are recited by the independent Claims 10, 18, and 27.

1. a plurality of audio metadata track extraction components in data communication with the output of the switch, wherein each audio metadata track extraction component provides an audio metadata track associated with speech

The Chow reference teaches one audio feature extractor which performs speech recognition on what is determined to be speech and outputs a string of words—text. In that sense, the Chow reference teaches only one extractor, namely a text extractor that receives the speech part of the audio waveform and transcribes it into text data. In Applicant’s embodiment of a method and engine, the equivalent text extraction is performed by the Speech Transcription Feature Extractor 620 as shown in Figure 10. In contrast to Chow, which only teaches a text extractor, however, Applicant’s claims include “a plurality of audio meta extraction components.” Figure 10 shows Speaker ID Feature Extractor 630 that is in data communication with the output of the Switch 610 and provides a metadata track associated with IDs of speakers 638 extracted from the speech. Figure 9 also illustrates that Applicant’s method and engine contemplates more than a text extractor. For example, the Feature Extractor Framework 510 includes several special feature extractors including KeyFrame Extractor 512, CC-text extractor 514, Audio Analysis Extractor 516, Speech-to-Text Extractor 518, and Speaker Extractor 520. In addition, Figure 10 illustrates that these special feature extractors are independent, modular units having their own dedicated dictionaries, not merely part of a text extractor (Speech Transcription Feature Extractor 620). Furthermore, Table 1 on page 12 lists various metadata tracks that are associated with the outputs of

these feature extractors, including Keyframe Track, CC-text Track, Audio-Class Track, Speech Track, Keyword Track, and Speaker ID Track.

2. an audio classification component controlling the audio signal switch according to whether the audio signal is classified as speech

This claim feature contains: the audio signal switch, such as switch 610, and the audio classification component, such as classification component 600, controlling the switch as shown in Figure 10. In one embodiment, the audio classification component 600 receives and classifies the audio signal into speech and non-speech categories. When the audio classification component detects the audio signal as "speech," it triggers audio signal switch 610, which then allows the digital audio signal 598 to pass into additional feature extractors which are capable of processing speech. In Chow, these two functions—classification and switching—are performed by a single component of VQ distortion processing block 303 which is described as follows:

VQ distortion processing block 303 performs sound classification to determine whether the sound waveform is speech or noise.... If VQ distortion processing block 303 determines that the sound waveform represents speech, then the sound waveform, in its processed state, is permitted to proceed to the speech recognition stage. On the other hand, if VQ distortion processing block 303 determines that the sound waveform represents noise, then the sound waveform is not permitted to proceed to the speech recognition stage. (Col. 7, ll. 30-40, Chow).

The combination of these two functions—speech/noise classification of the sound waveform and switching (passing/blocking) of the waveform depending on the outcome of the classification—in VQ distortion processing block this is clearly illustrated in Figure 3 where the VQ distortion processing block accepts digital audio sound waveform and outputs the waveform to the speech recognizer. There is no separate "audio classification component controlling" the VQ distortion processing block. Therefore, the Chow reference does not teach these two interacting but separate functional components.

Because the Chow reference does not teach or suggest the two claim features common to Claims 10, 18, and 27, Applicant respectfully requests that the independent Claims 10, 18, and 27 be allowed. Since each of Claims 13, 15, and 16 depends either directly or indirectly from Claim 10 and contains all the features of Claim 10, Applicant respectfully requests that those

Appl. No. : 10/067,550  
Filed : February 5, 2002

claims be allowed. Also, since each of Claims 20, 22, and 23 depends either directly or indirectly from Claim 18 and, contains all the features of Claim 20, Applicant respectfully requests that those claims be allowed. Furthermore, since each of Claims 22-23 depends either directly or indirectly from Claim 21 and, therefore, contains all the features of Claim 21, Applicant respectfully requests that those claims be allowed. Since each of Claims 30, 32, and 33 depends either directly or indirectly from Claim 27 and contains all the features of Claim 27, Applicant respectfully requests that those claims be allowed.

## **II. Discussion of the 35 U.S.C. § 103(a) Rejections**

Claims 1-3, 5-7, 9 and 24-26 were rejected under 35 U.S.C. § 103(a) as being obvious over Chang et al. (U.S. Patent No. 5,828,809) in view of Yoshio et al. (U.S. Patent No. 6,034,942). Claim 4 was rejected under 35 U.S.C. § 103(a) as being obvious over Chang et al. (U.S. Patent No. 5,828,809) in view of Yamashita (U.S. Patent No. 5,963,702). Claim 8 was rejected under 35 U.S.C. § 103(a) as being obvious over Chang et al. (U.S. Patent No. 5,828,809) in view of Yoshio et al. and further in view of Reichel et al. (U.S. Patent No. 5,701,153). Claims 11 and 28 were rejected under 35 U.S.C. § 103(a) as being obvious over Chow et al. in view of Miyamori et al. (U.S. Patent No. 5,677,994). Claims 12, 19, and 29 were rejected under 35 U.S.C. § 103(a) as being obvious over Chow et al. in view of Reichel et al. Claims 14, 21, 31, and 35 were rejected under 35 U.S.C. § 103(a) as being obvious over Chow et al. Claims 17 and 34 were rejected under 35 U.S.C. § 103(a) as being obvious over Chow et al. in view of Cruz et al. (U.S. Patent No. 5,613,032).

### **A. The Standard for Establishing a *Prima Facie* Case of Obviousness**

The standard for establishing a *prima facie* case of obviousness is set forth in M.P.E.P. §2143. "To establish a *prima facie* case of obviousness, three basic criteria must be met. First, there must be some suggestion or motivation, either in the references themselves or in the knowledge generally available to one of ordinary skill in the art, to modify the reference or to combine reference teachings. Second, there must be a reasonable expectation of success. Finally, the prior art reference (or references when combined) must teach or suggest all the claim

limitations. The teaching or suggestion to make the claimed combination and the reasonable expectation of success must both be found in the prior art, and not in Applicant's disclosure."

B. Discussion of Distinctions of Claims 1-3, 5-7, and 24-26 over Chang in view of Yoshio.

Applicant submits that neither Chang nor Yoshio teaches or suggests all the claim features of independent Claims 1 and 24. In particular, neither Chang nor Yoshio teaches the claim feature reciting *"extracting a plurality of audio metadata tracks from the audio information."*

The Chang reference is directed to video indexing system that automatically extracts indexing information from a video tape based on its audio and video components. Specifically, Chang describes extracting audio components from audio-video data for a video tape and spotting, from the audio component, a word that has a match in a database and creating an index around the time-location of the spotted word. The database contains audio-features of words that are desired to be spotted. As with the Chow reference discussed in Section I-B, the Chang reference does not teach special feature extractors and corresponding dedicated metadata tracks. Although the Chang reference describes speech recognition for recognizing particular words that are stored in a database, the Chang reference does not teach or suggest extracting a plurality of audio metadata tracks, each of which is dedicated for a special feature such as a keyword or an ID of a speaker.

While discussing the Chang reference, the Office Action states: "It is pertinent to point out that the examiner reads, for example, the data representing sounds, pointers (or indexes) to the locations, and the locations, where the specific words (e.g., keywords) are found, as examples of audio metadata tracks." (p. 6, the last sentence of the first continued paragraph). Applicant respectfully submits that data representing sounds and pointers (or indexes) to the locations and the locations where the specific words are found do not constitute "a plurality of audio metadata tracks" within the meaning of the phrase as used in Claims 1 and 24. Pointers to the locations and the locations where the specific words are found are merely means of addressing and are not audio metadata tracks. Applicant has described (on p. 2 of the specification) that metadata includes "information [ ] about the content of the video stream in real-time." As described at pages 9-10 of the specification, the information about the content of the stream (metadata) is shown in a time-based track representation, where each metadata track is a parcel of metadata spanning a time period and is extracted by a corresponding feature extractor. Furthermore, it is well known,

Appl. No. : 10/067,550  
Filed : February 5, 2002

and is recited in the assignee's other patents (e.g., U.S. Patent No. 6,877,134) that metadata is data describing some underlying data to be accessed. Therefore, it is submitted that Chang does not disclose "*extracting a plurality of audio metadata tracks from the audio information.*"

The Yoshio reference is directed to a recording method for recording information (video information, audio information, and the like) onto a high density information recording medium, and a reproducing method for reproducing the information from the information record medium wherein the video and audio data are divided on the basis of time codes. Yoshio does not teach or suggest speech feature extractors and corresponding dedicated metadata tracks.

Because neither the Chang reference nor the Yoshio reference teaches or suggests the claim feature reciting "extracting a plurality of audio metadata tracks from the audio information," Applicant submits that Claims 1 and 24 are not made obvious by combination of these references. Therefore, Applicant respectfully requests withdrawal of the rejection of Claims 1 and 24 under 35 U.S.C. §103(a) and requests those claims be allowed. Each of Claims 2, 3, 5, and 7 depends from and, therefore, contains all the features of Claim 1. Therefore, Applicant respectfully requests withdrawal of the rejections of Claims 2, 3, 5, and 7. Similarly, each of Claims 25 and 26 depends from and, therefore, contains all the features of Claim 24 and requests those claims be allowed. Therefore, Applicant respectfully requests withdrawal of the rejections of Claims 25 and 26 and requests those claims be allowed.

C. Discussion of Distinctions of Claims 12, 19, and 29 over Chow in view of Reichek.

Each of Claims 12, 19, and 29 depends directly from and, therefore, contains all the features of Claims 10, 18, and 27, respectively. Therefore, the 103(a) obviousness rejections of Claims 12, 19, and 29 over Chow in view of Reichek are not proper unless both references either individually or in combination teach or suggest all the claim features of Claim 10, 18, and 27, respectively. Applicant submits that neither Chang nor Reichek teaches or suggests all the claim features of independent Claims 10, 18, and 27. In particular, neither Chang nor Yoshio teaches the common claim feature reciting "*a plurality of audio metadata track extraction components in data communication with the output of the switch, wherein each audio metadata track extraction component provides an audio metadata track associated with speech.*"

This claim feature includes: multiple audio metadata track extraction components and multiple audio metadata tracks, each of which is dedicated to each of speech features extracted by the extraction components. Applicant's specification supports both multiple extractors as discussed in Section I-B; and multiple audio metadata tracks, including one metadata track dedicated to full text 628, other special-feature metadata tracks including one metadata track dedicated to keywords 648, and one metadata track dedicated to ID of speakers 638 as shown in Figure 10. As discussed in Section I-B above, the Chow reference does not teach or suggest separate metadata track extraction components dedicated to special speech features and, therefore, also does not teach separate metadata tracks dedicated to special speech features.

Applicant submits that Reichek also does not teach or suggest neither multiple extraction components nor multiple audio metadata tracks. The Reichek reference relates to a method and apparatus for inserting a time code into a textural representation of a speech part of a video data coming from a video tape recorder (VTR). The Reichek reference does not teach separate metadata tracks dedicated to special features. Instead, under Reichek, various extractable features including text and keywords are all lumped into a single metadata track, and there is no additional special-feature metadata track dedicated to a special speech feature such as a keyword or a speaker ID. Therefore, Reichek does not teach or suggest multiple audio metadata tracks.

For the reasons given above, neither Chow nor Reichek teaches or suggest the claim feature reciting "a plurality of audio metadata track extraction components in data communication with the output of the switch, wherein each audio metadata track extraction component provides an audio metadata track associated with speech," which are common to the independent Claims 10, 18, and 27 from which Claims 12, 19, and 29 depend. Therefore, Applicant respectfully requests withdrawal of the rejection of Claims 12, 19 and 29 under 35 U.S.C. §103(a) over Chow in view of Reichek and requests those claims to be allowed.

D. Discussion of Distinctions of Claims 11 and 28 over Chow in view of Miyamori.

Each of Claims 11 and 28 depends directly from and, therefore, contains all the features of Claims 10 and 27, respectively. Therefore, the 103(a) obviousness rejections of Claims 11 and 28 over Chow in view of Miyamori are not proper unless both references either individually or in combination teach or suggest all the claim features of Claim 10 and 27, respectively.

Applicant submits that neither Chow nor Miyamori teaches or suggests all the claim features of independent Claims 10 and 27. In particular, neither Chow nor Miyamori teaches the common claim feature reciting *“a plurality of audio metadata track extraction components in data communication with the output of the switch, wherein each audio metadata track extraction component provides an audio metadata track associated with speech.”*

As discussed in Section I-B and Section II-B above, the Chow reference does not teach or suggest this claim feature. The Miyamori reference is directed to method and apparatus for high-efficiency encoding and decoding of multi-channel data. Miyamori does not teach or suggest speech feature extractors and corresponding dedicated metadata tracks. Therefore, the Miyamori reference also does not teach or suggest this claim feature.

Because neither the Chow reference nor the Miyamori reference teaches or suggests the claim feature found in Claims 10 and 27 reciting *“a plurality of audio metadata track extraction components in data communication with the output of the switch, wherein each audio metadata track extraction component provides an audio metadata track associated with speech,”* Applicant submits that each of Claims 11 and 28, which depends from Claims 10 and 27, respectively, is not made obvious by combination of these references. Therefore, Applicant respectfully requests withdrawal of the rejections of Claims 11 and 28 under 35 U.S.C. §103(a) over Chow in view of Miyamori and requests the claims be allowed.

E. Discussion of Distinctions of Claims 14, 21, 31, and 35 over Chow.

Claim 14 depends from and, therefore, contains all the features of Claim 10. Therefore, the 103(a) obviousness rejection of Claim 14 over Chow is not proper unless Chow teaches or suggests all the claim features of Claim 10. Claim 21 depends from and, therefore, contains all the features of Claim 18. Therefore, the 103(a) obviousness rejection of Claim 21 over Chow is not proper unless Chow teaches or suggests all the claim features of Claim 18. Each of Claims 31 and 35 depends from and, therefore, contains all the features of Claim 27. Therefore, the 103(a) obviousness rejections of Claims 31 and 35 over Chow are not proper unless Chow teaches or suggests all the claim features of Claim 27. Applicant submits that Chow does not teach or suggest all the claim features of independent Claims 10, 18, and 27. In particular, as discussed in Section I-B and Section II-C, above, the Chow reference does not teaches the common claim



Appl. No. : 10/067,550  
Filed : February 5, 2002

feature reciting “a plurality of audio metadata track extraction components in data communication with the output of the switch, wherein each audio metadata track extraction component provides an audio metadata track associated with speech.”

Because the Chow references does not teach or suggest the claim feature commonly found in Claims 10, 18, and 27 reciting “a plurality of audio metadata track extraction components in data communication with the output of the switch, wherein each audio metadata track extraction component provides an audio metadata track associated with speech,” Applicant submits that each of Claims 14, 21, 31, and 35, which depends from Claims 10, 18, and 27, is not made obvious by Chow. Therefore, Applicant respectfully requests withdrawal of the rejections of Claims 14, 21, 31, and 35 under 35 U.S.C. §103(a) over Chow and requests the claims be allowed.

F. Discussion of Distinctions of Claims 17 and 34 over Chow in view of Cruz.

Each of Claims 17 and 34 depends directly from and, therefore, contains all the features of Claims 10 and 27, respectively. Therefore, the 103(a) obviousness rejections of Claims 17 and 34 over Chow in view of Cruz are not proper unless both references either individually or in combination teach or suggest all the claim features of Claim 10 and 27, respectively. Applicant submits that neither Chow nor Cruz teaches or suggests all the claim features of independent Claims 10 and 27. In particular, neither Chow nor Miyamori teaches the common claim feature reciting “a plurality of audio metadata track extraction components in data communication with the output of the switch, wherein each audio metadata track extraction component provides an audio metadata track associated with speech.”

As discussed in Section I-B and Section II-C above, the Chow reference does not teach or suggest this claim feature. The Cruz reference is directed to a digital camcorder as an audio signal source. Cruz does not teach or suggest speech feature extractors and corresponding dedicated metadata tracks. Therefore, the Cruz reference also does not teach or suggest this claim feature.

Because neither the Chow reference nor the Cruz reference teaches or suggests the claim feature found in Claims 10 and 27 reciting “a plurality of audio metadata track extraction components in data communication with the output of the switch, wherein each audio metadata track extraction component provides an audio metadata track associated with speech,” Applicant submits that each of Claims 17 and 34, which depends from Claims 10 and 27, respectively, is not

Appl. No. : 10/067,550  
Filed : February 5, 2002

made obvious by combination of these references. Therefore, Applicant respectfully requests withdrawal of the rejections of Claims 17 and 34 under 35 U.S.C. §103(a) over Chow in view of Cruz and requests the claims be allowed.

G. Discussion of Distinction of Claim 4 over Chang in view of Yamashita

Claim 4 depends directly from and, therefore, contains all the features of Claim 1. Therefore, the 103(a) obviousness rejection of Claim 4 over Chang in view of Yamashita is not proper unless both references either individually or in combination teach or suggest all the claim features of Claim 1. Applicant submits that neither Chang nor Yamashita teaches or suggests all the claim features of Claim 1. In particular, neither reference teaches or suggests the common claim feature reciting “*a plurality of audio metadata track extraction components in data communication with the output of the switch, wherein each audio metadata track extraction component provides an audio metadata track associated with speech.*”

For reasons given in Section II-B above, Chang does not teach or suggest this claim feature. The Yamashita reference is directed to using satellite broadcast tuners for receiving information. Yamashita does not teach or suggest speech feature extractors and corresponding dedicated metadata tracks. Therefore, Applicant submits that the Yamashita reference also does not teach or suggest this claim feature.

Because neither the Chang reference nor the Yamashita reference teaches or suggests the claim feature reciting “*a plurality of audio metadata track extraction components in data communication with the output of the switch, wherein each audio metadata track extraction component provides an audio metadata track associated with speech,*” Applicant submits that Claim 4 is not made obvious by combination of these references. Therefore, Applicant respectfully requests withdrawal of the rejection of Claim 4 under 35 U.S.C. §103(a) over Chang in view of Yamashita and requests the claim be allowed.

H. Discussion of Distinction of Claim 8 over Chang in view of Yoshio and further in view of Reichel

Claim 8 depends directly from and, therefore, contains all the features of Claim 1. Therefore, the 103(a) obviousness rejection of Claim 4 over Chang in view of Yoshio and further

**Appl. No. : 10/067,550**  
**Filed : February 5, 2002**

in view of Reichek is not proper unless all three references either individually or in combination teach or suggest all the claim features of Claim 1. Applicant submits that neither Chang nor Yoshio teaches or suggests all the claim features of Claim 1. In particular, neither reference teaches or suggests the common claim feature reciting *"a plurality of audio metadata track extraction components in data communication with the output of the switch, wherein each audio metadata track extraction component provides an audio metadata track associated with speech."*

For reasons given in Section II-B above, Chang does not teach or suggest this claim feature. The Yoshio reference is directed to recording method for recording information (video information, audio information, and the like) onto high density information recording medium, and reproducing method for reproducing the information from the information record medium wherein the video and audio data are divided on the basis of time codes. Yoshio does not teach or suggest speech feature extractors and corresponding dedicated metadata tracks. Therefore, the Yoshio reference also does not teach or suggest this claim feature. For reasons given in Section II-C above, Reichek also does not teach this claim feature.

Because neither the Chang reference nor the Yoshio reference nor the Reichek reference teaches or suggests the claim feature reciting *"a plurality of audio metadata track extraction components in data communication with the output of the switch, wherein each audio metadata track extraction component provides an audio metadata track associated with speech,"* Applicant submits that Claim 8 is not made obvious by combination of these references. Therefore, Applicant respectfully requests withdrawal of the rejection of Claim 8 under 35 U.S.C. §103(a) over Chang in view of Yoshio and further in view of Reichek and requests the claim be allowed.

### **III. New Claims**

Applicant has added new dependent Claims 36-40. Claims 46-40 are supported at least by Figure 6.

### **IV. Conclusion**

In light of the above, reconsideration and withdrawal of the outstanding rejections are specifically requested. In view of the foregoing remarks, Applicant respectfully submits that the claims of the above-identified application are in condition for allowance. However, if the

Appl. No. : 10/067,550  
Filed : February 5, 2002

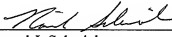
Examiner finds any impediment to allowing all claims that can be resolved by telephone, the Examiner is respectfully requested to call the undersigned.

Please charge any additional fees, including any fees for additional extension of time, or credit overpayment to Deposit Account No. 11-1410.

Respectfully submitted,

KNOBBE, MARTENS, OLSON & BEAR, LLP

Dated: January 17, 2007

By:   
Raimond J. Salenieks  
Registration No. 37,924  
Agent of Record  
Customer No. 20,995  
(619) 235-8550

3224591  
121806